World Journal of Economics and Business Research
ISSN: 2960-0081
DOI: https://doi.org/10.61784/wjebr3036

AUTOMATIC PRICING AND REPLENISHMENT MODEL OF VEGETABLE PRODUCTS BASED ON MULTIPLE LINEAR REGRESSION

JingYao Wang*, YiXuan Yang, LinSen Song
School of Insurance and Economics, University of International Business and Economics, Beijing 100029, China.
Corresponding Author: JingYao Wang, Email: m17705380176@163.com

Abstract: In fresh supermarkets, the shelf life of vegetable products is often relatively short, and the quality of the product always turns worse over time. In order to achieve the minimum loss and maximum profit, it is crucial for supermarkets to make reasonable replenishment and pricing decisions for the sold vegetable products.Traditional pricing and replenishment strategies are inefficient, making it difficult to accurately control costs and consuming a significant amount of manpower.This article comprehensively processes and analyzes the sales data of a supermarket, establishes a sales price relationship daily replenishment pricing model, and successfully predicts the daily replenishment volume and pricing of the target date. It is hoped to provide a systematic daily replenishment and pricing strategy for supermarket in the sales of fresh vegetables and vegetables, and help supermarket achieve maximum self-interest.
Keywords: Daily replenishment pricing model; Demand price theorem; Multiple regression; Binding item

1 INTRODUCTION

Fresh products are an indispensable necessity in daily life, and people have a great demand for them. Due to the difficulty in preserving fresh products, people tend to purchase them in small quantities multiple times, resulting in some fresh supermarkets that are close to home and at moderate prices being favored by people[1]. Among fresh products, vegetable products play a major role in product sales. The shelf life of vegetable products is generally short, and they often have an immediate nature during the sales process - most varieties cannot be sold the next day if they are not sold on the same day. In order to meet consumers' pursuit of product diversity, supermarkets often sell a wide variety of vegetables, and the purchase transaction time for vegetables is usually from 3:00 to 4:00 in the morning[2-3]. Therefore, merchants often need to make replenishment decisions for each vegetable category on the day without knowing the specific individual product and purchase price.

The traditional pricing method for vegetable products is a subjective cost plus pricing method based on factors such as the color, selling time, and quality of vegetable products by supermarket operators. This type of method not only consumes a lot of energy for supermarket operators, but also makes it difficult for supermarket operators to achieve precise cost control due to unclear purchase prices of goods[4]. The various types of vegetable items make precise pricing even more difficult, ultimately making it difficult for supermarket operators to maximize their own interests.

This article attempts to establish a daily replenishment pricing model based on a multiple linear regression model, taking into account multiple influencing factors such as purchase volume, loss rate, and product type. To help supermarket operators achieve precise pricing of fresh vegetables and vegetables while saving manpower and resources, and improve customer satisfaction; Helping consumers have a diverse selection of product types, obtain reasonable product prices, and ultimately support the high-quality development of fresh agricultural product sales terminals.

2 PROBLEM DESCRIPTION

This article establishes an appropriate mathematical model, under the assumption that supermarkets make replenishment plans based on categories, and first analyzes the relationship between the total sales volume of each vegetable category and cost plus pricing. Based on this, and with the goal of maximizing the revenue of supermarkets, the daily replenishment volume of each category of vegetables in the next week is given.

Using the processed data to analyze the trend of changes in individual product pricing and sales volume, it was determined that there is a linear relationship between pricing and sales. Therefore, a multiple linear regression model was used to solve the problem. We will use data from total sales volume and cost plus pricing to perform multiple linear fitting on the sales volume of various vegetable categories, with total sales volume as the dependent variable and cost and other related variables as independent variables. We will fit the model and predict the total sales volume for the next week,determining the total replenishment amount based on the relationship between total sales and replenishment amount (Total sales=(1-Loss rate) · Replenishment amount).And finally, determine the total replenishment amount and pricing strategy based on the cost plus pricing method.Considering the limited sales space of vegetable products, further development of replenishment plans for individual items is proposed.

3 MODEL ASSUMPTIONS

Considering that the sales data of supermarkets are greatly influenced by external environmental factors, such as the impact of holidays on the customer flow of supermarkets; The behavior of supermarket operators themselves can also greatly affect sales data, such as supermarket promotional activities, etc; The behavior of different supermarket operators can also have a mutual impact on their sales data; There is also a mutual influence relationship between prices of different items and categories. To study the sales data of supermarkets for three years, due to the limited practical research scope of sales data, and to ensure the rigor and reliability of the model and final results, the following assumptions are made:
(1)The sales data provided by the supermarket for three years is accurate and complete, and there are no duplicate or incorrect records.
(2)The sales data of supermarket in the short term remains stable and will not be affected by certain unexpected factors such as natural disasters.
(3)Market conditions remain stable, the behavior of competitors in supermarkets remains stable, and their pricing strategies and sales plans will not experience significant fluctuations; Consumer preferences and purchasing power remain relatively stable.
(4)The historical sales data of the supermarket can predict the future sales of the supermarket, and the historical sales data can serve as a reference for predicting future sales data.
(5)The sales space of the supermarket will not undergo significant changes in a short period of time, and the sales space of the supermarket is constant for a certain period of time.
(6) Pricing will to some extent affect sales volume. Pricing based on market rules is negatively correlated with sales volume.
(7) Unless there is clear data showing correlation, the correlation between individual items and categories is not considered, and the sales volume of each item and category are treated as independent data.

4 SYMBOL DESCRIPTION

Symbols that used in the article are in the Table 1 below.

Table 1 Symbol Description

Symbol	Description
Q_{ij}	Predict the sales volume of category j on day i,or the sales volume of individual item j on day i.
h_{ij_1}, h_{ij_2}, h_{ij_3}, ……, h_{ij_m}	All independent variables related to category j, including pricing,time, correlation between categories, loss rate, etc.There are m independent variables in total.
b_{j_1}, b_{j_2}, b_{j_3}, ……, b_{j_m}.	The model parameters of all independent variables related to i in category j, including pricing, time, correlation between categories, loss rate, etc., totaling m independent variables.
a_j	Model parameters independent of j variety and i
ϵ	Error term of the model
P_{ij}	Pricing of j categories/items on day i
c_{ij}	Cost of category/item j on day i
r_{ij}	Profit margin of category/item j on day i
R_{ij}	Profit of category/item j on day i
γ_j	j category/item loss rate
W_{ij}	Replenishment quantity for category/item j on day i
n	The total number of saleable items

5 ESTABLISHMENT OF A DAILY REPLENISHMENT PRICING MODEL FOR SALES PRICE RELATIONSHIP

5.1 The relationship between total sales volume and cost plus pricing

Before solving this problem, we first introduce the demand theorem in microeconomics. The demand theorem is one of the fundamental theorems in economics, which states that the demand for a commodity varies inversely with its price, while other conditions remain unchanged. Simply put, the higher the price of a product, the fewer people buy it; The

lower the price, the more people will buy. There is a certain correlation between sales volume and cost pricing. Therefore, here we choose the multiple regression prediction method to solve the problem [5-8].
The multiple regression prediction method is a statistical method used to determine the relationship between one or more independent variables and the dependent variable. The following is the form of the relationship:
Set the sales volume of the product as Qij. The basic form of the relationship is:

$$Q_{ij} = a_j + b_{j_1}h_{ij_1} + b_{j_2}h_{ij_2} + b_{j_3}h_{ij_3} \ldots\ldots + b_{j_m}h_{ij_m} + \epsilon \tag{1}$$

·Q_{ij} is the dependent variable (the sales volume we want to predict)

·h_{ij_1}, h_{ij_2}, h_{ij_3}, ..., h_{ij_m} are independent variables (influencing factors, including pricing, time, correlation between categories, loss rate, etc.)

·aj, b_{j_1}, b_{j_2}, b_{j_3}, ······, b_{j_m}, are parameters

·ϵ is error term

To facilitate the solution and avoid over fitting, the expression is simplified.

$$Q_{ij} = a_j + b_{ij}P_{ij} \tag{2}$$

Therefore, only the key value of sales volume Pij is selected for expression. Use the least squares method to determine the various parameters of the model.

5.2 Establishment and Solution of Optimization Model

Due to limitations in the sales space of supermarkets, it is necessary to choose the best sales combination in order to maximize profits. For optimization problems, choose to establish a sales price relationship daily replenishment pricing model for solving. The pricing strategy can be determined by the relationship between sales volume and cost pricing obtained in 5.1, so only sales volume prediction is needed here.
Considering that sales volume will be influenced by multiple factors such as time and wholesale price, a multivariate fitting model will be established using Python. Using the sales volume of vegetable categories as the dependent variable and wholesale prices as the independent variable.
The cost pricing method is based on the cost and profit of goods as support during the pricing process, therefore, it is set:Q_{ij} represents the sales volume of category j on day i, P_{ij} represents the pricing of category j on day i, c_{ij} represents the cost of category j on day i, r_{ij} represents the profit margin of category j on day i, and R_{ij} represents the profit margin of category j on day i. A_j and b_j are constants.

$$Q_{ij} = a_j + b_jP_{ij} \tag{3}$$

The pricing of category j on day i can be expressed as:

$$P_{ij} = (1 + r_{ij})c_{ij} \tag{4}$$

The profit of category j on day i can be expressed as:

$$R_{ij} = Q_{ij}r_{ij} \tag{5}$$

The total profit of goods is expressed as:

$$R_{total} = \sum_{i=1}^{n}\sum_{j=1}^{m} R_{ij} \tag{6}$$

For constraint conditions, the equation constraint is the equation relationship between sales volume and pricing; The inequality constraint is that all variables are non negative, and the sales volume should be within the range of the historical maximum and minimum, namely:

$$Q_{min} \leq Q \leq Q_{max} \tag{7}$$

In summary, the optimization model is:

$$max\ R_{total} = \sum_{i=1}^{n}\sum_{j=1}^{m} R_{ij} \tag{8}$$

$$s.t.\begin{cases} P_{ij} = a_j + b_j Q_{ij} \\ P_{ij} = (1 + r_{ij}) c_{ij} \\ R_{ij} = Q_{ij} r_{ij} \\ P_{ij}，Q_{ij}，R_{ij} \geq 0 \\ R_{ij} \in R \\ \text{i} = 1，2，......\text{n} \\ \text{j} = 1,2,......\text{m} \\ \text{（Representing different vegetable varieties）} \end{cases} \tag{9}$$

At the same time, provide the relationship between the loss rate and replenishment quantity, assuming the loss rate is γ j. If the replenishment quantity is W_{ij}, then:

$$Q_{ij} = （1 - \gamma_j）W_{ij} \tag{10}$$

5.3Establishment of a Model Considering Constrained Items

The revenue of supermarkets is influenced by various factors such as the type of replenishment items, the minimum display quantity of each item, and the amount of product replenishment. Therefore, it is necessary to determine the optimal replenishment quantity and pricing strategy for the optimal product combination under the constraint of the number of item types to maximize the revenue of supermarkets. For optimization problems with multiple influencing factors, multiple regression prediction methods are used, Establish a constrained single item sales price relationship daily replenishment pricing model to solve the problem. In order to establish an optimization model, R_j is measured by r_j and constrained to be within the [0,0.2] interval. The optimization model is used to solve and obtain the optimal solution of the final r_j. On the basis of the previous model, the corresponding functional relations and constraints can be obtained, and the final solution function is as follows:

(1) The relationship between sales volume Q_{ij} and pricing P_{ij} is:

$$Q_{ij} = a_j + b_j P_{ij} \tag{11}$$

(2) The corresponding relationship between pricing $P_{ij,}$cost C_{ij}, and interest r_{ij} rate is:

$$P_{ij} = (1 + r_{ij}) C_{ij} \tag{12}$$

(3) The relationship between profit R_{ij},cost C_{ij}, price P_{ij} and sales Q_{ij} is：

$$R_{ij} = (P_{ij} - C_{ij}) Q_{ij} \tag{13}$$

(4) The corresponding relationship between salesQ_j,loss rateγ_j,and replenishmentW_j is：

$$Q_j = (1 - \gamma_j) W_j \tag{14}$$

(5) The newly added constraint is that the total number of items that can be sold should be between n1 and n2, and the display quantity of each item should be at least x kilograms, taking n as the total number of items.

$$\text{n1} \leq \text{n} \leq \text{n2}$$

$$W_j \geq 2.5$$

$$R_{total} = max \sum_{i=1}^{j} R_j \tag{15}$$

$$s.t.\begin{cases} Q_j = a_j + b_j P_j \\ P_j = (1 + r_j) C_j \\ R_j = (P_j - C_j) Q_j \\ Q_j = (1 - \gamma_j) W_j \\ P_j, Q_j \geq 0 \\ W_j \geq x \\ R_j \in R \\ j = 1,2,3 \ldots\ldots n 1 \leq n \leq n2 \end{cases} \tag{16}$$

6 ALGORITHM DESIGN

6.1 Algorithm Design of Total Sales and Cost Plus Pricing Model

(1) Merge sales data and wholesale price data: First, group the sales data by date and item code, and calculate the daily total sales volume.
(2) Check the column names of wholesale price data, redefine column names to match sales data, convert sales date columns to date format, and merge the data again.
(3) Define X (independent variable) and y (dependent variable).
(4) Divide the training and testing sets, fit the multiple linear regression model, and calculate the R ^ 2 value of the model to evaluate the degree of fit of the model.

6.2 Algorithm Design for Optimizing Models

(1) Group sales data by category and statistically describe the sales volume of vegetables in different categories. To make it more intuitive, draw a histogram of the distribution of sales volume.
(2) Draw a box diagram.
(3) Calculate the correlation coefficient between categories, and based on the statistical description of sales volume and the category correlation coefficient matrix, ultimately establish an optimization model for vegetable sales volume.

6.3 Establishment of a Model Considering Constrained Items

(1) Recalculate the slope and intercept of the linear regression model.
(2) Extracting Slope and Intercept from a Linear Model.
(3) Perform the optimization process again.
(4) Ensure that the display quantity of each item is at least x kilograms.
(5) Obtain the optimized cost profit margin and establish an optimization model that considers constrained individual products. Calculate the expected sales volume and pricing strategy for each product based on the optimized cost-profit margin.

7 MODEL ANALYSIS AND SOLUTION

7.1 Daily Replenishment Volume and Pricing Strategy for Each Category of Vegetables in the Next Week

This article establishes an appropriate mathematical model, under the assumption that supermarkets make replenishment plans based on categories, and first analyzes the relationship between the total sales volume of each vegetable category and cost plus pricing. Based on this, and with the goal of maximizing the revenue of supermarkets, the daily replenishment volume and pricing strategy for each category of vegetables in the next week, from July 1 to 7, 2023, are given.
Set the sales volume of the product as Q_{ij} and only select the key value of sales volume P_{ij} to represent it. Using the least squares method, calculate the various parameters of the model, and the results are shown in the table below. Table 2 to Table 7 represent the parameters of cauliflower, eggplant, mosaic, chili, edible fungi, and aquatic rhizomes, respectively.

Table 2 The Result of Cauliflower

	B	Standard Error
a_j	20.94	2.19
b_j	-15.68	3.47

Table 3 The Result of Eggplant

	B	Standard Error

a_j	13.27	0.96
b_j	-8.33	1.60

Table 4 The Result of Mosaic

	B	Standard Error
a_j	12.08	0.60
b_j	-0.59	0.69

Table 5 The Result of Chili

	B	Standard Error
a_j	9.51	0.91
bj	-3.62	1.91

Table 6 The Result of Edible Fungi

	B	Standard Error
a_j	10.00	1.02
b_j	-3.83	1.83

Table 7 The Result of Aquatic Rhizomes

	B	Standard Error
a_j	13.83	0.82
b_j	-10.99	1.72

Based on historical data, predict the sales volume and pricing strategy of each individual product from July 1st to July 7th, 2023. The pricing strategy can be obtained from the relationship between the sales volume and the cost pricing obtained from the previous model, so only the sales volume can be predicted here.

Considering that sales volume is influenced by multiple factors such as time and wholesale price, combined with existing data from the previous three years and the conclusion drawn from the first question, a multivariate fitting model is established using Python.

Using the sales volume of vegetable categories as the dependent variable and wholesale prices as the independent variable. Firstly, the aquatic rhizome category is selected as an example for the above analysis, and then the predicted sales results of the aquatic rhizome category in the next week are given, as shown in the Table 8 below.

Table 8 The predicted sales results of the aquatic rhizome category in the next week

Time	Sales volume（kg）
2023/7/1	18.95
2023/7/2	16.32
2023/7/3	17.17
2023/7/4	16.73
2023/7/5	18.61
2023/7/6	18.40
2023/7/7	18.49

The cost pricing method is supported by the cost and profit of the goods during the pricing process, and determines the pricing of the goods through this relationship. For constraint conditions, the equation constraint is the equation relationship between sales volume and pricing; The inequality constraint is that all variables are non negative, and the sales volume should be within the range of historical maximum and minimum quantities. Based on the established model, the daily replenishment volume and pricing strategy for each vegetable category in the next week are obtained, as shown in the Table 9 and Table 10 below.

Table 9 The Replenishment Volume and Pricing Strategy for each Vegetable Category

	Cauliflower		Mosaic		Chili	
Date	Total daily replenishment（kg）	Pricing strategy(%)	Total daily replenishment（kg）	Pricing strategy (%)	Total daily replenishment（kg）	Pricing strategy (%)
2023-07-01	53.35	50.06%	211.46	33.25%	75.14	75.14%
2023-07-02	20.12	25.25%	260.46	27.43%	108.96	108.95%
2023-07-03	41.15	50.15%	261.42	42.88%	109.55	109.55%
2023-07-04	39.27	87.36%	164.15	20.38%	43.09	43.09%
2023-07-05	50.92	41.12%	262.48	91.23%	110.03	110.02%
2023-07-06	31.45	39.27%	260.47	38.74%	108.93	108.92%
2023-07-07	29.18	52.17%	203.12	55.48%	70.09	70.08%

Table 10 The Replenishment Volume and Pricing Strategy for each Vegetable Category

	Eggplant		Edible fungi		Aquatic rhizomes	
Date	Total daily replenishment（kg）	Pricing strategy (%)	Total daily replenishment（kg）	Pricing strategy (%)	Total daily replenishment（kg）	Pricing strategy (%)
2023-07-01	9.79	96.00%	65.20	33.06%	18.95	84.08%
2023-07-02	8.73	41.80%	66.91	100.03%	16.32	44.21%
2023-07-03	5.62	64.84%	31.85	107.01%	17.17	88.24%
2023-07-04	10.84	105.51%	49.31	86.00%	16.73	32.74%
2023-07-05	3.30	95.34%	46.76	92.50%	18.61	52.91%
2023-07-06	6.62	109.11%	58.33	91.30%	18.40	33.92%
2023-07-07	6.15	84.10%	61.97	62.41%	18.49	38.11%

7.2 Replenishment and Pricing Strategy for Vegetable Items on July 1st

Considering the limited sales space of vegetable products, it is required to provide a further replenishment plan for each item. Specifically, this article requires a replenishment quantity and pricing strategy for each item on July 1 based on the available varieties from June 24 to 30, 2023, and provides conditional constraints. The constraints are as follows:
(1) Control the total number of items that could be sold between 27 and 33;
(2) The minimum order quantity for each item should meet the display quantity of 2.5kg;
(3) Try to meet the market's demand for various types of vegetable products;
(4) The supermarket has the highest revenue.
On the basis of the established model that considers the constraint of individual products, during the process of substituting data for solving, it was found that the data for June 2023 was partially missing. Considering the accuracy of the results, the data from May 2023 was used as a substitute. Based on expected profits, 30 items with the highest expected profits were selected and a model was established using Python. The final results are shown in the Table 11. below.

Table 11 Replenishment and Pricing Strategy for Vegetable Items on July 1st

Number	Replenishment	Price
01	226.00	3.95
02	150.00	3.39
03	149.00	3.32
04	133.00	1.85
05	99.60	3.70
06	93.00	4.56
07	87.90	11.43
08	79.00	3.77
09	79.00	2.15
10	76.30	4.88
11	73.00	5.47
12	70.00	4.10
13	62.50	3.52
14	62.00	2.45
15	49.00	3.78
16	49.00	2.99
17	47.90	7.85
18	45.00	3.01
19	42.20	8.05
20	41.50	5.25
21	34.30	3.54
22	32.40	18.58
23	31.50	5.44
24	29.50	9.08
25	28.20	21.15

8 CONCLUSION

(1) By solving the model, we found that compared from a category perspective: the total sales of flowers and leaves are the largest and significantly higher than other types; The total sales volume of eggplants is the smallest; The total sales volume of chili peppers and fungi used is relatively close; Finally, the total sales of cauliflower and aquatic rhizomes are basically the same. The pricing strategies of the same category on different dates often have significant fluctuations. Taking aquatic rhizome vegetables as an example, the standard deviation of the pricing strategy from July 1 to 7, 2023 is 23.38037, indicating that there is significant fluctuation in the pricing of aquatic rhizome vegetables within a week.
(2) This article analyzes sales data and establishes appropriate mathematical models to develop a daily replenishment plan and pricing strategy for the supermarket, in order to maximize the revenue of the supermarket and meet the

demand of consumers for various types of vegetables. Therefore, the purpose and significance of this article are from the perspectives of both enterprises and consumers.

(3) From the perspective of enterprises, developing a reasonable daily replenishment plan and pricing strategy can improve sales efficiency, better grasp the sales situation and market demand of vegetable products, timely replenish goods, avoid inventory backlog or shortage, control costs, and ultimately achieve benign pricing of products and improve profit margins. At the same time, better meet customer needs, ensure sufficient supply of goods and reasonable prices, and improve customer shopping experience and satisfaction.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] He X, Gao J. The operating mechanism and reconstruction countermeasures of China's fresh agricultural product circulation system under the new development pattern. Commercial & Economic Research, 2023, (14): 5-8.

[2] Guo Z, Li K, Qi H. A decision model for the latest product development of enterprises. Journal of Hebei University (Natural Science Edition), 2022(02): 108-111.

[3] Yinxu. Research on joint ordering and forwarding strategies for three locations. Ximen University, 2017.

[4] Guo B, Li Y. Market competition and pricing analysis of green products. Market Modernization, 2023(22): 9-12.

[5] Xie Y. Regression Analysis. Beijing: Social Science Academic Press, 2013.

[6] Menzel. Sales Forecast: Method System Management. Beijing: Posts and Telecommunications Press, 2004.

[7] Luo P. Pricing Strategy (Fourth Edition). Shanghai: Shanghai University of Finance and Economics Press, 2019.

[8] Zhao C, Li N, Jiang Z, et al. Parametric optimization of ambient and cryogenic loop heat pipes using multiple linear regression method. International Journal of Refrigeration, 2024, 161:145-163.